AD 742090

Best Available Copy

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Department of Statistics Stanford University, Calif	2a. REPORT SECURITY CLASSIFICATION 2b. GROUP

3. REPORT TITLE

ON INTERVAL ESTIMATION AND SIMULTANEOUS SELECTION OF ORDERED LOCATION OR SCALE PARAMETERS

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

TECHNICAL REPORT

5. AUTHOR(S) (Last name, first name, initial)

RIZVI, M. Haseeb and SAXENA, K. M. Lal

6. REPORT DATE May 5, 1972	7a. TOTAL NO. OF PAGES 11	7b. NO. OF REFS 3
8a. CONTRACT OR GRANT NO. N00014-67-A-0112-0053 b. PROJECT NO. NR-042-267 c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) No. 194 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)	

10. AVAILABILITY/LIMITATION NOTICES

Distribution of this document is unlimited

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY Office of Naval Research Stat & Probability Program Code 436 Arlington, VA.

13. ABSTRACT

A formulation is given and a procedure is proposed for constructing a confidence interval for a certain ordered (location or scale) parameter and for simultaneously selecting all populations having parameters equal or larger than this ordered parameter with a preassigned minimal probability. The well-known indifference-zone formulation of the ranking problem is obtained as a special case as is the problem of interval estimation of an ordered parameter.

DD FORM 1 JAN 64 1473

UNCLASSIFIED
Security Classification

UNCLASSIFIED

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Interval Estimation						
Simultaneous Selection						
Ordered Parameter						
Indifference Zone of Ranking Formulation						

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2a. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2b. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year, or month, year. If more than one date appears on the report, use date of publication.

7a. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7b. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8a. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8b, 8c, & 8d. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9a. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9b. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through ___________."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through ___________."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through ___________."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring *(paying for)* the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as *(TS)*, *(S)*, *(C)*, or *(U)*.

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, roles, and weights is optional.

DD FORM 1 JAN 64 1473 (BACK)

Unclassified

Security Classification

ON INTERVAL ESTIMATION AND SIMULTANEOUS SELECTION OF ORDERED LOCATION OR SCALE PARAMETERS

by

M. Haseeb Rizvi and K. M. Lal Saxena

TECHNICAL REPORT NO. 194

May 5, 1972

PREPARED UNDER CONTRACT N00014-67-A-0112-0053

(NR-042-267)

OFFICE OF NAVAL RESEARCH

Herbert Solomon, Project Director

DEPARTMENT OF STATISTICS

STANFORD UNIVERSITY

STANFORD, CALIFORNIA

ON INTERVAL ESTIMATION AND SIMULTANEOUS

SELECTION OF ORDERED LOCATION OR SCALE PARAMETERS

By M. Haseeb Rizvi and K. M. Lal Saxena
Stanford University and University of Nebraska

1. Introduction and Formulation of the Problem .

Procedures for selection of a certain number of populations with larger parameters from a collection of several populations have been studied extensively in the past two decades; see, for example, Barr and Rizvi [1] for a simple exposition. Recently Saxena and Tong [2] and Saxena [3] have considered confidence intervals for the largest parameter. The present paper attempts to combine these two requirements simultaneously in a single formulation. The problem of interest is to construct a confidence interval for a certain ordered parameter and simultaneously select all populations having parameters equal or larger than this ordered parameter, with a preassigned minimal probability whenever parameters lie in a specified subspace. A procedure R is proposed to solve this problem, and its performance in terms of probability requirement being satisfied is evaluated.

Consider $k(\geq 1)$ populations $\pi_i(i=1,\ldots,k)$ with absolutely continuous distribution function (df) $F(.;\ \theta_i)$ of Y_i on the real line with real parameter θ_i and let $f(.;\ \theta_i)$ be the corresponding density. Let $\theta_{[1]} \leq \theta_{[2]} \leq \cdots \leq \theta_{[k]}$ denote the ordered values of the components of $\underset{\sim}{\theta} = (\theta_1, \theta_2, \ldots, \theta_k) \in \Omega$. For $1 \le t \le k$, we require a procedure R that selects all π_i with $\theta_i \ge \theta_{[k-t+1]} = \theta$ (say) and simultaneously gives an interval I such that $\theta \in I$. Denote by CS the (correct) selection of all π_i with $\theta_{[i]}, i = k-t+1, \ldots, k$ and by CD (correct decision) the inclusion of θ in I and let $P(\underset{\sim}{\theta})$ denote $\Pr\{CS \cap CD | R\}$. Then the procedure R, for some preassigned constant γ, $1/\binom{k}{t} < \gamma < 1$, is more specifically required to satisfy

$$(1.1) \qquad \inf_{\Omega(\psi)} P(\underset{\sim}{\theta}) \ge \gamma ,$$

where $\Omega(\psi) = \{\underset{\sim}{\theta} \in \Omega: \theta_{[k-t]} \le \psi(\theta_{[k-t+1]})\}$ and ψ is a given function on the real line such that $\psi(x) \le x$.

2. Main Results on $P(\underset{\sim}{\theta})$ for Proposed R.

Proposed Procedure R.

Rank $Y_1, Y_2, \ldots, Y_k$, breaking ties (if any) with suitable randomization, and let $Y_{[i]}$ be the i^{th} smallest Y_i. Consider two suitably chosen continuous increasing functions h_1 and h_2 (with inverses g_2 and g_1 respectively). Construct the random interval $I_0 = (h_1(Y_{[k-t+1]}), h_2(Y_{[k-t+1]}))$. Then assert that $\theta \in I_0$ and that the π_j's corresponding to $Y_{[j]}(j = k-t+1, \ldots, k)$ have parameters $\theta_j \ge \theta$.

We shall presently investigate the infinum of $P(\underset{\sim}{\theta})$ over $\Omega(\psi)$ for the above R and later determine conditions so that R satisfies (1.1). We have

$$(2.1)\qquad P(\underset{\sim}{\theta}) = \sum_{j=k-t+1}^{k} \int_{g_1(\theta)}^{g_2(\theta)} \prod_{r=1}^{k-t} F(y;\ \theta_{[r]}) \prod_{\substack{s=k-t+1 \\ s\neq j}}^{k} (1 - F(y;\ \theta_{[s]}))$$

$$dF(y;\ \theta_{[j]})\ .$$

An obvious proposition follows.

Proposition 1.

A sufficient condition that $P(\underset{\sim}{\theta})$ be a nonincreasing function of $\theta_{[1]}, \ldots, \theta_{[k-t]}$ is that the df's $F(\cdot,\ \theta_i)$, $i = 1, \ldots, k$ be stochastically ordered.

Location Parameter Case.

Let $F(y,\ \theta_i) = F(y - \theta_i)$, $\psi(\theta) = \theta - \delta$, $g_1(\theta) = \theta - a$, $g_2(\theta) = \theta + b$, where $\delta \geq 0$ and a and b with $a+b > 0$ are given constants; $\Omega(\psi)$ will now be denoted by $\Omega(\delta)$. Clearly, (2.1) implies

Proposition 2.

For $t = 1$,

$$(2.2)\qquad \inf_{\Omega(\delta)} P(\underset{\sim}{\theta}) = \int_{-a}^{b} F^{k-1}(y+\delta)\,dF(y)\ .$$

Theorem 1.

Suppose $f(y-\theta_1)$ has a monotone likelihood ratio (m.l.r.) in y for θ_1 and constants a and b are chosen such that $a+b > 0$ and

$$F(-a) + F(b) \geq 1 \quad . \tag{2.3}$$

Then, for $1 < t \leq k$,

$$\inf_{\Omega(\delta)} P(\underset{\sim}{\theta}) = P(\underset{\sim}{\theta}_o) = t \int_{-a}^{b} F^{k-t}(y+\delta)[1-F(y)]^{t-1} dF(y), \tag{2.4}$$

where $\underset{\sim}{\theta}_o$ has first $(k-t)$ components equal to $(\theta-\delta)$ and the last t components equal to θ, θ being any arbitrary value of $\theta_{[k-t+1]}$.

Proof.

Since $f(y, \theta_1)$ has an m.l.r., Proposition 1 implies that $P(\underset{\sim}{\theta})$ is minimized over Ω_1 by setting $\theta_{[1]} = \ldots = \theta_{[k-t]} = \theta-\delta$, where Ω_1 is the subset of $\Omega(\delta)$ for which $\theta_{[k-t+1]}, \ldots, \theta_{[k]}$ are held fixed. Letting $\theta-\theta_{[j]} = \delta_j (\leq 0)$, $j=k-t+2, \ldots, k$, we obtain from (2.1) after some simplification,

$$(2.5) \qquad \inf_{\Omega_1} P(\underset{\sim}{\theta}) = F^{k-t}(\delta-a)[1-F(-a)] \prod_{j=k-t+2}^{k} [1-F(\delta_j-a)]$$

$$+F^{k-t}(\delta+b)[1-F(b)] \prod_{j=k-t+2}^{k} [1-F(\delta_j+b)]$$

$$+(k-t)\int_{\delta-a}^{\delta+b} F^{k-t-1}(y)[1-F(y-\delta)] \prod_{j=k-t+2}^{k}$$

$$[1-F(y-\delta+\delta_j)]\,dF(y)$$

$$= H(\underset{\sim}{\delta}), \text{ say,}$$

where $\underset{\sim}{\delta} = (\delta_{k-t+2},\ldots,\delta_k)$. Since $H(\underset{\sim}{\delta})$ is a symmetric function of its arguments, minimization of $H(\underset{\sim}{\delta})$ over $\{\underset{\sim}{\delta}: \delta_{k-t+2} \geq \cdots \geq \delta_k\}$ is equivalent to its minimization over $\Omega_2 = \{\underset{\sim}{\delta}: \delta_{k-t+2} \leq 0,\ldots,\delta_k \leq 0\}$. For some j, fix $\delta_{k-t+2},\ldots,\delta_{j-1},\delta_{j+1},\ldots,\delta_k$ and consider $(\partial/\partial\delta_j)H(\underset{\sim}{\delta})$. Observe that the m.l.r condition implies that $f(\delta_j-a)/f(\delta_j+b)$ and $f(y+\delta_j-\delta)/f(\delta_j+b)$ are increasing functions of δ_j, for all $y \in (\delta-a, \delta+b)$. Arguing in a similar manner as in Saxena [3], we conclude that $(\partial/\partial\delta_j)H(\underset{\sim}{\delta})$ has at most one change of sign from positive to negative and consequently $\inf_{\delta_j} H(\underset{\sim}{\delta})$ is either at $\delta_j = 0$ or at $\delta_j = -\infty$. This conclusion is valid for every other j. Therefore infinum of $H(\underset{\sim}{\delta})$ over Ω_2 is achieved when a certain number r of

δ_j's are zero and the rest equal to $-\infty$; denote this infimum by $G(r)$. Then (2.5) after integration by parts gives

$$(2.6)\qquad G(r) = (r+1)\int_{-a}^{b} F^{k-1}(y+\delta)[1-F(y)]^{r}\,dF(y)\ ,$$

where $r\epsilon\{0,1,\ldots,t-1\}$. Now it follows from the lemma given below that $G(t-1) \leq G(r)$ for all $r = 0,1,\ldots,t-1$. Consequently,

$$(2.7)\qquad \inf_{\Omega(\delta)} F(\underset{\sim}{\theta}) = \inf_{\Omega_2} H(\underset{\sim}{\delta}) = G(t-1),$$

which proves the theorem.

Lemma.

A sufficient condition for $G(r)$, given by (2.6), to be nonincreasing in r is that a and b are such that $F(-a)+F(b) \geq 1$.

Proof.

Consider the following density function

$$(2.8)\qquad h(y;\,r) = [C(r)]^{-1}(r+1)[1-F(y)]^{r}f(y),\ -a < y < b,\quad \text{where}$$

$$(2.9)\qquad C(r) = [1-F(-a)]^{r+1} - [1-F(b)]^{r+1}\ .$$

With E_r denoting the expectation with respect to (2.8), we can write (2.6) as

(2.10) $$G(r) = C(r)E_r\{F^{k-t}(Y+\delta)\} \; .$$

Since $h(y;\ r)/h(y;\ s)$ is an increasing function of y for $r > s$,

(2.11) $$E_r\{F^{k-t}(Y+\delta)\} \geq E_s\{F^{k-t}(Y+\delta)\} \; .$$

Therefore, $G(r) \geq G(s)$ if $C(r) \geq C(s)$ which is implied by the condition of the lemma.

Scale Parameter Case.

Let $F(y;\ \theta_i) = F(y/\theta_i), y > 0, \theta_i \geq 0, \psi(\theta) = \rho\theta, g_1(\theta) = \theta/a, g_2(\theta) = \theta/b$, where ρ, a, b are given constants such that $0 < \rho \leq 1,\ 0 \leq b < a$; $\Omega(\psi)$ will now be denoted by $\Omega(\rho)$. We now state the following results, the proofs for which are readily constructed along the lines of the ones given for the location parameter case.

Proposition 3.

For $t = 1$,

(2.12) $$\inf_{\Omega(\rho)} P(\underset{\sim}{\theta}) = \int_{1/a}^{1/b} F^{k-1}(y/\rho)\,dF(y) \; .$$

Theorem 2.

Suppose $f(y/\theta_i)$ has an m.l.r. in y for θ_i and constants a and b are chosen such that

$$F(1/a) + F(1/b) \geq 1 . \tag{2.13}$$

Then, for $1 < t \leq k$,

$$\inf_{\Omega(\rho)} P(\underset{\sim}{\theta}) = P(\underset{\sim}{\theta}_o) = t \int_{1/a}^{1/b} F^{k-t}(y/\rho)[1 - F(y)]^{t-1} dF(y) , \tag{2.14}$$

where $\underset{\sim}{\theta}_o$ now has first $(k - t)$ components equal to $\rho\theta$ and the last t components equal to θ, θ being any arbitrary value of $\theta_{[k-t+1]}$.

3. Some Other Formulations as Special Cases.

A noteworthy feature of the present formulation is that the $\Pr\{CS \cap CD|R\}$ is minimized at $\underset{\sim}{\theta}_o$, defined after (2.4) in the location parameter case and after (2.14) in the scale parameter case. This $\underset{\sim}{\theta}_o$ is also the "least favorable configuration" for the indifference zone formulation of the ranking problem (see [1]) as well as for the confidence interval formulation (see [2] and [3]). Thus the present work includes the ranking formulation as a special case; with $a = b = \infty$ in the location parameter case and $a = \infty$, $b = 0$ in the scale parameter case, $\Pr\{CS \cap CD|R\}$ equals $\Pr\{CS|R\}$ and (2.2) and (2.4) reduce to (7) of [1] and (2.12) and (2.14) to (10) of [1].

The present formulation also includes the confidence interval formulation for the largest or the smallest parameter as a special case. For $t = k$ we have $\theta = \theta_{[1]}$, $\Omega(\delta) \equiv \Omega$ and $\Pr\{CS \cap CD|R\}$ equals $\Pr\{CD|R\}$.

Thus for the smallest location parameter (2.4) yields

$$(3.1) \qquad \inf_{\Omega} P(\underset{\sim}{\theta}) = [1 - F(-a)]^k - [1 - F(b)]^k ,$$

provided $F(-a) + F(b) \leq 1$. Letting $Y_i' = -Y_i$ and $\theta_i' = -\theta_i (i = 1, \ldots$ we obtain for the largest location parameter,

$$(3.2) \qquad P(\underset{\sim}{\theta}) = Pr(Y_{[k]} - b < \theta_{[k]} < Y_{[k]} + a)$$

$$= Pr(\theta'_{[1]} - b < Y'_{[1]} < \theta'_{[1]} + a)$$

and, therefore in view of (3.1),

$$(3.3) \qquad \inf_{\Omega} P(\underset{\sim}{\theta}) = F^k(b) - F^k(-a) ,$$

provided $F(-a) + F(b) \leq 1$. Note that with $a = b = d$ and $F \equiv G_n$, (3.3) reduces to (4) of [2]. Similar discussion holds for the scale parameter case and the related result of [3].

4. Applications.

Consider k populations π_i with real parameters θ_i, $i = 1, \ldots, k$. Considerations of invariance under the permutation of the indices of the k populations suggest taking random samples of a common size n from each population. Let Y_i be a function of the sufficient statistic (when it exists) for θ_i and let its df be $F_n(\cdot; \theta_i)$; this df plays the role of $F(.; \theta_i)$ of the above discussion. In order that the procedure R of Section 2 satisfy (1.1), the smallest n should be determined such that (2.2) or (2.4) ((2.12) or (2.14)) is at least as large as the preassigned constant γ. Such a solution exists if Y_i's are consistent. As an illustration let π_i be $N(\theta_i, 1)$, $i = 1,\ldots,k$. Then Y_i's are sample means based on random samples each of size n and $F_n(y,\theta_i) = \Phi(n^{1/2}(y - \theta_i))$ where $\Phi(.)$ is the standard normal df. Now (2.4) gives

$$(4.1) \qquad \inf_{\Omega(\delta)} P(\underset{\sim}{\theta}) = t \int_{-an^{1/2}}^{bn^{1/2}} \Phi^{k-t}(y + n^{1/2}\delta)[1 - \Phi(y)]^{t-1} d\Phi(y) ,$$

where $b \geq a$. The right side of (4.1) tends to unity for $b \geq a > 0$, so that there is a unique n satisfying (1.1).

5. Concluding Remarks.

It should be noted that if $\delta = 0$ ($\rho = 1$) then the integral (2.4) (integral (2.14)) can be evaluated with the help of the incomplete beta function tables and the tables of the df F; in addition if $a = b = \infty$ ($a = \infty$, $b = 0$), $\inf_{\Omega} P(\underset{\sim}{\theta}) = 1/\binom{k}{t}$.

In this formulation of interval estimation and simultaneous selection, the upper confidence bound for $\theta_{[k-t+1]}$ can be obtained by taking $b = \infty (b = 0)$ and some finite a, satisfying conditions of Theorem 1 (Theorem 2). However, the conditions of Theorem 1 (Theorem 2) do not permit the construction of the lower confidence bound for $\theta_{[k-t+1]}$ except in the trivial case $a = \infty$, $b = \infty$ ($a = \infty$, $b = 0$).

REFERENCES

[1] Barr, D. R. and Rizvi, M. H. (1966). An introduction to ranking and selection procedures. Jour. Amer. Stat. Assoc. **61**, 640-646.

[2] Saxena, K. M. Lal and Tong, Y. L. (1969). Interval estimation of the largest mean of k normal populations with known variances. Jour. Amer. Stat. Assoc. **64**, 296-299.

[3] Saxena, K. M. Lal (1971). Interval estimation of the largest variance of k normal populations. Jour. Amer. Stat. Assoc. **66**, to appear.